A3223-5

Vol. 62 No. 8 August 1987

JOURNAL OF Medical Education

PUBLICATION OF THE ASSOCIATION OF AMERICAN MEDICAL COLLEGES

Pilot Study Using 'Dangerous Answers' as Scoring Technique on Certifying Examinations
Henry J. Mankin, John S. Lloyd, and Richard J. Rovinelli

Validity of Scoring 'Dangerous Answers' on a Written Certification Examination
Stephen Slogoff and Francis P. Hughes

Faculty Development Fellowship Programs in Family Medicine
Carole J. Bland, Maurice A. Hitchcock, William A. Anderson, and Frank T. Stritter

Relationship of Residents' Emotional Problems, Coping Behaviors, and Gender
Elizabeth H. Young

Substance Use, Eating Behaviors, and Social Impairment of Medical Students
David B. Herzog, Jonathan F. Borus, Paul Hamburg, Ingrid L. Ott, and Adriane Concus

Journal of Medical Education

Volume 62 • Number 8 • August 1987

Pilot Study Using 'Dangerous Answers' as Scoring Technique on Certifying Examinations—Henry J. Mankin, John S. Lloyd, and Richard J. Rovinelli 621

Validity of Scoring 'Dangerous Answers' on a Written Certification Examination—Stephen Slogoff and Francis P. Hughes 625

Faculty Development Fellowship Programs in Family Medicine—Carole J. Bland, Maurice A. Hitchcock, William A. Anderson, and Frank T. Stritter 632

Relationship of Residents' Emotional Problems, Coping Behaviors, and Gender—Elizabeth H. Young 642

Substance Use, Eating Behaviors, and Social Impairment of Medical Students—David B. Herzog, Jonathan F. Borus, Paul Hamburg, Ingrid L. Ott, and Adriane Concus 651

Psychiatric Interviews and Psychological Tests as Predictors of Medical Students' Success—C. Knight Aldrich 658

Using Patient Instructors To Teach Behavioral Counseling Skills—Jeffrey C. Levenkron, Philip Greenland, and Nancy Bowley 665

Faculty Assessments of Case Reports of Medical Students—P. J. McLeod 673

Communications: Academic Predictors of Minority Students' Acceptance into Medical School—Carmen Montecinos and John Pohlmann 678

The Intellectual Caliber of Medical Students of the 1980s and Past Decades—James V. Lupo, Robert E. Mitchell, Thomas G. Grandy, and Charlene Erskine Combs 680

Gender Differences in Predictors of Performance in Medical Training—E. Virginia Calkins, Louise M. Arnold, and T. Lee Willoughby 682

Physician Manpower in Utah—Daniel O. Clegg, Judith P. Short, and Cecil O. Samuelson, Jr. 685

A One-Day Communication Workshop for Internal Medicine Residents—John F. Christensen, Wendy Levinson, Judy L. Colligan, Patrick M. Dunn, Stephen R. Jones, and Alan Morgenstern 687

A Comparison of Two Formats for Teaching Critical Reading Skills in a Medical Journal Club—Mark Linzer, Elizabeth R. DeLong, and Kenneth H. Hupart 690

Comparison of Computer-Assisted and Seminar Learning of Electrocardiogram Interpretation by Third-Year Students—Ruth-Marie E. Fincher, Abdulla M. Abdulla, Marandapalli R. Sridharan, Warren H. Gullen, John S. Edelsberg, and John S. Henke 693

Training Pediatrics Residents To Use Clinical Observation Scales—David A. Bergman, Robert H. Pantell, and Ross DeHovitz 695

Editorials: Minority Admissions: Quo Vadis? 698

Exploring Wrong Answers 700

Bibliography 702

Letter to the Editor 706

Officers of the Association and Members of the Executive Council, 1986–87

CHAIRMAN: Edward J. Stemmler, *University of Pennsylvania*
CHAIRMAN-ELECT: John W. Colloton, *University of Iowa Hospitals and Clinics*
IMMEDIATE PAST CHAIRMAN: Virginia V. Weldon, *Washington University*
PRESIDENT: Robert G. Petersdorf, *AAMC, Washington, D.C.*

COUNCIL OF DEANS
Louis J. Kettel, *University of Arizona**
William T. Butler, *Baylor College of Medicine*
D. Kay Clawson, *University of Kansas*
Robert S. Daniels, *LSU at New Orleans*
William B. Deal, *University of Florida*
John Naughton, *SUNY at Buffalo*
Richard S. Ross, *Johns Hopkins University*
Henry P. Russe, *Rush Medical College*
Hibbard E. Williams, *UC at Davis*

COUNCIL OF ACADEMIC SOCIETIES
Frank G. Moody, *University of Texas, Houston**
David H. Cohen, *Northwestern University*
William F. Ganong, *UC at San Francisco*
Douglas E. Kelly, *University of Southern California*

COUNCIL OF TEACHING HOSPITALS
Spencer Foreman, *Montefiore Medical Center**
J. Robert Buchanan, *Massachusetts General Hospital*
Gary Gambuti, *St. Luke's-Roosevelt Hospital Center*
C. Thomas Smith, *Yale-New Haven Hospital*

ORGANIZATION OF STUDENT REPRESENTATIVES
Vicki Darrow, *University of Washington**
Kimberly Dunn, *University of Texas, Houston*

DISTINGUISHED SERVICE MEMBER
Edward N. Brandt, Jr., *University of Maryland*

* Chairman of respective organization.

Validity of Scoring 'Dangerous Answers' On a Written Certification Examination

Stephen Slogoff, M.D., and Francis P. Hughes, Ph.D.

Abstract—The American Board of Anesthesiology (ABA) investigated the possibility that selection of a disproportionately high number of incorrect "dangerous answers" by candidates who would otherwise pass the board's written certification examination might relate to dangerous clinical practice by the candidates. In that case, scoring the examination for dangerous answers would provide an alternate scoring method. Twenty-nine multiple-choice questions in the 1983 written examination were identified by a panel of judges as containing incorrect answers with "truly, not potentially" dangerous consequences. The 1,036 candidates who passed the 1983 examination selected an average of 1.6 dangerous answers, and the 1,413 who failed the examination selected an average of 3.4 ($p < .001$). Ninety-two candidates who passed and who selected four or more dangerous answers were tracked through the certification process to learn whether labeling them as "smart but dangerous" was justified. Reports of their clinical competence from their residency programs did not relate dangerous behavior for any of these candidates. The 86 of these 92 candidates who subsequently took the ABA oral examination had pass rates that were identical to those of all candidates in 1983 and that were completely independent of the candidates' absolute and relative frequency of choosing dangerous answers. It is concluded that selection of dangerous answers in a multiple-choice test is the result of lack of information rather than a purposeful action and implementation of the scoring system for dangerous answers would likely be unjustified and unfairly punitive.

In 1984 the Committee on Study of Evaluation Procedures (COSEP) of the American Board of Medical Specialties reported to its member boards during its annual meeting the results of a preliminary study of candidates who selected incorrect responses that could be considered dangerous in answering multiple-choice questions on written certifying examinations. Specifically, a retrospective evaluation of examinations given by four member specialty boards was performed, and four distinct groups of candidates were identified in regard to their choice of "dangerous answers." There were two large groups: one consisted of passing candidates who selected few dangerous answers, and the other consisted of failing candidates who selected many dangerous answers. There were two small groups:

The study reported here was conducted by the American Board of Anesthesiology.

Dr. Slogoff is clinical professor of anesthesiology, University of Texas Health Science Center at Houston, and a director of the American Board of Anesthesiology; Dr. Hughes is executive secretary of the board.

For related editorial, see page 700.

one consisted of failing candidates who selected few dangerous answers and the other of those who passed but selected a disproportionately high number of dangerous answers.

In their presentation, the COSEP investigators speculated that this last group, the "smart but dangerous" candidates, might constitute a truly hazardous group of practitioners and might be identified by scoring dangerous answers separately in addition to grading the entire examination. The examination questions with dangerous answers would serve as a special subset of the examination. No further analyses were performed to determine whether the clinical patterns of this group justified labeling the group as "smart but dangerous." No published reports have been found in the literature that deal with validation of this grading methodology.

The American Board of Anesthesiology (ABA), which did not participate in the COSEP study, requires completion of three separate processes for certification as a consultant: receipt of a certificate of clinical competence from the candidate's residency training program, satisfactory performance on a written examination that primarily evaluates factual knowledge and relatively direct clinical judgment, and, no less than one year later, satisfactory performance on an oral examination that tests more sophisticated and complex aspects of clinical judgment and practice. Stimulated by the preliminary COSEP report, the ABA performed a similar pilot study in 1985 using its 1982 written examination. The 1982 examination was chosen to permit immediate retrospective tracking of the group of "smart but dangerous" candidates through the other two aspects of the certifying process. Although about 10 percent of those who passed the written examination were designated as in this dangerous group based on their selection of a higher number of dangerous incorrect responses, neither the clinical competence committee reports nor oral examination data on these candidates indicated any deviant performance. The present report, based on data from the 1983 ABA written examination, is a detailed attempt to validate the hypothesis that selection of a disproportionately high number of dangerous answers predicts dangerous clinical thinking or behavior.

Methods

The 1983 ABA written examination contained 350 questions, of which 175 were single-answer multiple-choice questions. Of these 175, a total of 34 were selected for review because each possibly contained at least one dangerous response and that response had been selected by at least one passing candidate. These were independently examined by the eight members of the ABA examination committee, who were instructed to identify responses that, if selected, would have "truly, not potentially" dangerous consequences. The answers that were correct were not identified for the judges. For purposes of this study, a response was considered dangerous if it was judged so by at least four of the eight judges. An example of a question considered in the exercise is:

> A patient with aortic stenosis has chest pain while waiting for surgery. The EKG shows sinus tachycardia at a rate of 110/min. Blood pressure is 140/90 mmHg. The best therapeutic measure to relieve the chest pain is administration of
>
> A. nitroglycerin sublingually.
> B. propranolol.
> C. diazepam.
> D. supplemental oxygen.
> E. sodium nitroprusside by infusion.

The correct answer is "B," and all eight judges independently perceived response "E" to have truly dangerous consequences if selected and enacted.

The final subset of questions examined was composed of 29 questions with 43 incorrect, dangerous responses. The median number of votes for a dangerous answer in the subset was six, and the mean was 5.6. Two judges each perceived the correct answers to two different questions to be dangerous; neither question was included in the subset.

The written certifying examinations of the 2,449 ABA candidates in 1983 were rescored for performance on the subset of dangerous answers. For those who selected at least four dangerous answers, the ratio of such answers to all incorrect responses of the 29 subset questions and the ratio of such answers to all incorrect responses to the total 350 questions were calculated separately. The residency clinical competence reports for these candidates were then examined to identify comments relating to unsafe practice or behavior. For those candidates who selected at least four dangerous answers and passed the written examination, their oral examination performance was reviewed for both success and examiners' subjective comments. The performance of American medical graduates (AMGs) and foreign medical graduates (FMGs) were separated and compared.

Selection of four or more dangerous answers was arbitrarily selected to define the potentially dangerous group because this frequency permitted tracking of about 100 candidates (about 10 percent of those who passed the written examination) through the entire certification process. Continuous data were statistically analyzed by *t*-test for unpaired data. Means are reported with standard deviations.

Results

Of the 2,449 ABA candidates who took the 1983 examination, 1,036 passed and selected an average of 1.6 (S.D. = 0.3; range 0–7) of the 29 dangerous answers, while the 1,413 who failed selected 3.4 (S.D. = 0.4; range 0–10) dangerous answers ($p < .001$). To focus the study on the candidates who selected dangerous answers with unusual frequency, only those 723 who selected four or more were further evaluated. In this group, 92 passing candidates chose an average of 4.3 (S.D. = 0.8; range 4–7) dangerous answers and 631 failing candidates selected an average of 5.1 (S.D. = 0.3; range = 4–10; $p < .05$). Table 1 shows these data.

In the 29-question subset, the mean ratio of dangerous answers to all incorrect answers was 0.42 (S.D. = 0.08) for the 92 passing candidates who selected at least four dangerous answers and 0.31 (S.D. = 0.04) for the 631 failing candidates who selected four or more ($p < .01$), as indicated in Table 1. A similar relationship was found when the ratios of dangerous answers to incorrect responses for the entire 350 questions were calculated. Dangerous answers always comprised a larger fraction of incorrect responses for the passing than failing candidates in comparisons at the same frequency of such selection, because the passing candidates had fewer incorrect responses.

Since the ABA requires all candidates in its examination system to have an overall satisfactory rating on a clinical competence committee report covering their final residency training period, the final rating for each candidate must of necessity be satisfactory. However, awarding unsatisfactory ratings in individual categories of competence—which consist of knowledge, judgment, clinical skills, and character traits—is permitted while still giving a satisfactory rating for overall clinical competence. On the final reports only one of the 92 passing candidates received one unsatisfactory rating (in knowledge), and none of the failing candidates was judged unsatisfactory in any category.

TABLE 1

Percentages of 92 Passing and 631 Failing Candidates Who Chose Four or More Dangerous Answers to 29 Multiple-Choice Questions Containing Dangerous Answers and the Ratio of These Candidates' Dangerous Answers to the Total Number of Incorrect Answers for These 29 Questions on the 1983 Written Certification Examination of the American Board of Anesthesiology*

No. of Dangerous Answers	Passing Candidates		Failing Candidates	
	Percent	Ratio	Percent	Ratio
4	73	0.35	42	0.27
5	16	0.41	26	0.31
6	8	0.45	16	0.35
7	3	0.57	9	0.38
8	0	—	4	0.41
9	0	—	2	0.46
10	0	—	1	0.48

* Dangerous answers were judged independently by eight specialists as having "truly, not potentially" dangerous consequences. A total of 2,449 candidates took the 1983 examination, and 1,036 passed.

Eighty-six of the 92 candidates who had passed the written examination and selected four or more dangerous answers had taken their oral examination by the time of the study. Of the 50 AMGs, 28 (58 percent) passed; and eight (22 percent) of the 36 FMGs passed. These pass rates are comparable with the 64 percent of all AMGs and 31 percent of all FMGs in the total sample. The differences between the candidates from the potentially dangerous group and the overall candidate pool were not significant.

All three of the candidates who chose seven dangerous answers passed the oral examination, as did one of the five who chose six, five of the 15 who chose five, and 27 of the 63 who chose four dangerous answers.

The mean ratios of dangerous answers to incorrect answers for the dangerous answers subset was 0.37 (S.D. = 0.13) for the 36 candidates who passed the oral examination and 0.36 (S.D. = 0.11) for the 50 who failed, as indicated in Table 2. Fourteen of the 23 (61 percent) AMGs with ratios less than 0.35 passed, as did 14 of the 27 (62 percent) with ratios the same as or greater than 0.35. For the FMGs, these results were 16 percent and 29 percent, respectively. The differences were not significant.

Summary comments by each of the four oral examiners for these 86 candidates were reviewed for references to unsafe practice or judgment. One FMG who had selected four dangerous answers among 13 incorrect responses in the dangerous answer subset (ratio 0.31) failed the oral examination and was noted by one examiner to be "slow, hesitant, vacuous in some important clinical areas. . . . I think he is dangerous." Another FMG who had chosen four dangerous answers (ratio 0.36) and had failed the oral examination was reported by one examiner to be "potentially dangerous." A failing FMG who selected six dangerous answers (ratio 0.46) was noted by one examiner to have "superficial knowledge. He goes through the motions clinically but underlying actions are weak. . . . He also did several things that were dangerous." None of the remaining 341 summary comments about the 86 candidates contained references to safety.

TABLE 2

Numbers of 50 American Medical Graduates (AMGs) and 36 Foreign Medical Graduates (FMGs) Who Selected Four or More Dangerous Answers to 29 Multiple-Choice Questions with Dangerous Answers on the 1983 Written Certification Examination of the American Board of Anesthesiology (ABA) and the Percentages Who Passed the ABA Oral Examination at Their First Opportunity, by the Ratio of the Dangerous Answers to All Incorrect Answers to the 29 Questions on the Written Examination*

Ratios	AMGs		FMGs	
	No.	Percent	No.	Percent
<0.30	3	33	8	0
0.30–0.34	20	65	11	27
0.35–0.39	6	67	6	17
0.40–0.44	11	36	6	33
0.45–0.49	2	50	2	0
≥0.50	8	62	3	33

* Dangerous answers were judged independently by eight specialists as having "truly, not potentially" dangerous consequences. A total of 2,449 candidates took the 1983 written examination, and 1,036 passed. Of those who passed, a total of 92 had selected four or more dangerous answers; of these, 86 had taken the oral examination by the time of the study.

Discussion

The primary justification for using a secondary scoring method such as dangerous answer scoring in addition to traditional correct answer scoring when grading written certifying examinations is to prevent candidates from passing who deserve to fail but would otherwise be successful on the test. The preliminary COSEP report suggested this rationale might apply to the use of the dangerous answer scoring to identify the small group described as smart but dangerous. Were the COSEP hypothesis valid, failure to grade for dangerous answers would permit this group of physicians to acquire board certification yet practice medicine dangerously. It is unlikely that sufficient information about subsequent practice patterns of such examinees could be obtained in a short enough period of time to substantiate the hypothesis.

The ABA oral examination by design and construction is believed to provide a more sensitive and direct assessment of current practice patterns and medical judgment than does the written examination. It seemed appropriate, therefore, to test the dangerous answer scoring system in the written examination against performance in the oral examination. Performance of a group of 92 (9 percent) candidates who passed the written examination and selected at least four dangerous answers was analyzed separately from that of the 944 passing examinees who selected fewer than four. Eighty-six of the 92 took the oral examination on their first opportunity and performed without significant difference from the oral examinees who had selected less than four dangerous answers, when compared as an entire subgroup or separated into AMG and FMG cohorts. Of more relevance, success on the oral examination was not related to either the absolute number of dangerous answers selected or the relative frequency of their selection. Assignment of "dangerous" behavior to candidates in

three of 344 subjective assessments of the 86 oral examinees by their examiners is not excessive in the ABA examination process. A review of comments made about another 86 candidates with the same pass rate in the oral examination showed a similar frequency of references to dangerous thinking and behavior. However, the grading format for the oral examination does not specifically request the examiner to comment on dangerous practices by candidates. If it did, unsafe or dangerous practices might be noted more often for all candidates.

The ratings of the potentially dangerous candidates on the residency clinical competence reports were not different from those for failing candidates with similar numbers of dangerous answers and, parenthetically, from those for all candidates in the examination sample. The absence of criticism in such reports is likely due to the natural attrition of poorly performing residents prior to completion of postgraduate education and to the perception of residency program directors and their clinical competence committees that these reports describe a minimum standard of performance for entrance into the examination system. Truly critical assessment of residents' performance appears to be an internal function of the program.

The design of the present study can be criticized. The candidates were not told prior to testing that incorrect answers and, specifically, incorrect dangerous answers would have negative consequences. Nevertheless, the exercise was intended to detect candidates who believed their dangerous answers were correct. Pursuing only those with a disproportionately high frequency of dangerous answers would likely reduce the risk of including examinees who occasionally selected a dangerous answer by guessing. Further, there were relatively few questions with dangerous answers in the examination (29 of 175 multiple-choice type). This low number of questions directly resulted from three aspects of the exercise: (*a*) although many questions had dangerous answers, retrospective selection permitted using only those that had been chosen by at least one passing candidate, a necessary characteristic for the subsequent analysis; (*b*) the judges were instructed to select only answers which were truly dangerous and not just potentially dangerous; and (*c*) at least half of the judges had to identify the answer as truly dangerous. The authors believe the requirement for all three increased the validity of the subsequent observations.

The higher ratios of dangerous answers to incorrect responses in the entire test and in the dangerous answer subset for passing candidates, at all frequencies of selecting dangerous answers, simply reflect the relatively fewer number of incorrect responses by the passing examinees rather than membership in a uniquely identifiable group. Passing candidates chose less than half the number (1.6) of dangerous answers chosen by failing candidates (3.4). This difference also supports the conclusion that selection of a dangerous answer results primarily from the opportunity to select a dangerous answer by virtue of not knowing the correct answer; that is, selection of a dangerous answer reflects the absence of knowledge rather than a purposeful act. It also appears unlikely to the authors that attempts to identify the "smart but dangerous" candidate by creating dangerous responses prospectively rather than selecting them retrospectively would alter the outcome of this exercise.

In summary, the authors examined the validity of dangerous answer scoring within the limits of assessing concurrent

rather than future practice patterns and judgment. No evidence from residency reports on clinical competence or from evaluations of oral examiners was found to relate dangerous clinical practice to frequent selection of dangerous responses. Therefore, implementation of alternate scoring by the dangerous answer format would be unnecessarily punitive and unwarranted.

Relationship of Residents' Emotional Problems, Coping Behaviors, and Gender

Elizabeth H. Young, Dr.P.H.

Abstract—Family medicine residents were studied to test the hypotheses that a higher incidence and greater severity of personal or emotional problems are associated with a higher incidence of maladaptive coping behaviors among women than men residents. Self-report data were collected from 466 residents in six states on personal or emotional problems and on alcohol use, cigarette smoking, weight, diet, and exercise. The study provides evidence that the women residents experienced a higher incidence and greater severity of personal or emotional problems than men and that these problems were associated with an increased percentage of women than men who had used alcohol during the preceding year, had used alcohol daily, had increased use of alcohol over the two years before the survey, and had had perceptions of being overweight and were on calorie restricted diets. The gender differences in cigarette smoking and frequency of exercise were not significant. The author concludes that residency programs need to provide flexible programs that would better accommodate women's needs and to help residents develop healthy strategies for dealing with stress.

Studies that address the psychological well-being of family medicine residents suggest that residents experience considerable occupational stress and personal strain during the training program and feel they lose control over their environment and have inadequate social-support systems (1–4). Sources of occupational stress are frequently defined in terms of role conflicts such as role overload, role insufficiency, role ambiguity, role boundary, and role responsibility (1). Fatigue, lack of leisure time, isolation from life outside the hospital, and ambivalence about their career choice are problems associated with the physical environment in which they work (4). The hypothesis that the residency program itself is a pronounced source of stress because it produces in the residents feelings of loss of control over their environment is supported in one of the studies (3). All of these studies focused on differences in sources of stress and the effects of stress on residents at different levels of training or in comparison with faculty members, but the studies did not explore gender differences.

The effects of residency training on the personal and professional lives of residents during the first year of residency training were the focus of other studies (5–7). The incidence of major depression was found to be significantly higher than expected among first-year residents (5). Risk factors for major depression were

Dr. Young is an assistant professor, Department of Family Medicine, University of Texas Medical Branch, Galveston, Texas.

identified as having a history of major depression, being female, and being unmarried. Anger was reported to increase during the year, and sleep deprivation was constant throughout. Time pressures, fatigue, lack of self confidence, and lack of social support were also identified as problems more acutely felt during the first year than in subsequent years of residency training (6).

Women family medicine residents reported higher levels of stress than their male colleagues in reports of a Canadian study (7, 8). Problems in combining a personal and professional life were most frequently cited as sources of stress. Internal medicine program directors reported that leaves of absence granted to residents because of emotional impairment occurred twice as frequently among women as men; however, less than 2 percent of the residents were reported to require this type of leave (9).

Measures of psychological functioning among students at one medical school indicated that there were no gender differences in the students' psychological adjustment on admission to medical school but that by mid-year women students were found to receive less family support for their decision to pursue a medical career and to experience more internal conflict about their decision (10). Other studies indicated that women medical students found it harder to make friends because of responsibilities outside their training than did the men students, felt more isolated from other people than men students, and had greater difficulty meeting people with similar interests (11, 12). Studies of practicing physicians indicated that fewer women physicians were married and fewer had children than did men physicians (13, 14).

The evidence that women may experience more occupational stress and personal strain than men during their formative years in medicine deserves to be studied more thoroughly among larger and more varied populations than previous studies have done. Questions of major interest that were not adequately answered in previous studies were the focus of the investigation reported here. The following hypotheses were tested in the present study. Are there significant gender differences in the incidence of personal or emotional problems among family medicine residents? Are there significant gender differences in behaviors that are considered maladaptive coping mechanisms, such as cigarette smoking, increased alcohol use, excessive calorie intake, and infrequent exercise? Are these maladaptive behaviors related to the incidence and severity of the problems?

Method

The study population was composed of family medicine residents who were receiving their training in the six states of the south-central region of the United States—Missouri, Kansas, Oklahoma, Arkansas, Louisiana, and Texas. Questionnaires were sent in 1985 to the 781 family medicine residents in university-administered or university-affiliated training programs in these six states. These programs were categorized as three types: community-based, university-affiliated; community-based, university-administered; university-based.

The survey items on behavior were adapted from the questions used in the National Survey of Personal Health Practices and Consequences (15) conducted by the U.S. Department of Health and Human Services, because these questions had been extensively field-tested. The questions were designed to elicit information regarding the incidence and severity of personal or emotional problems during the preceding year; health prac-

tices related to smoking, alcohol use, exercise, and diet; and demographic characteristics.

The presence of personal or emotional problems was assessed by a bivariate response of yes or no; the items that dealt with frequency of health practices were measured on 3-, 4-, or 5-point Likert-type scales.

The respondents were assured that all responses would be treated as confidential information. A three-stage follow-up method of personal correspondence (that is, addressed specifically to each individual) was used to optimize the response rate. The original questionnaire was accompanied by a personal letter that explained the purpose of the study and included an addressed, stamped envelope for return. One week later a follow-up postcard served to thank those who had replied promptly and to offer a reminder to those who had not. A replacement questionnaire and a second personal letter were sent three weeks after the original to those who had not replied. A final replacement questionnaire and letter were sent seven weeks after the original mailing to nonrespondents.

Results

A total of 466 family medicine residents responded to the survey (60 percent). There were no significant differences in response rate by gender, postgraduate year of training, type of training program, or state. The gender distribution of the respondents very closely approximated the distribution in the study population: 78 percent of the responders were men, and 80 percent of the study population were men. These differences were not statistically significant.

Some demographic characteristics of the respondents by gender are shown in Table 1. There were no significant differences in age between men and women residents. The mean age of respondents was 29.4 years for men and 29.2 years for women. Although there were differences in the total number of women residents at each level of training, the percentage

TABLE 1

Number and Percentage of 466 Responding Residents in 33 Family Medicine Residency Programs in Six States in the South Central Region by Postgraduate Year of Training and Type of Program, 1985*

Year and Program	Men		Women	
	No.	Percent	No.	Percent
Postgraduate year				
1	118	32.4	37	35.9
2	124	34.1	29	28.2
3	121	33.5	37	35.9
Type of Program				
Community-based, university-affiliated	133	36.6	23	22.0
Community-based, university-administered	102	29.2	35	34.0
University-based, university-administered	128	35.2†	45	44.0†

* The six states were Missouri, Kansas, Oklahoma, Arkansas, Louisiana, and Texas.

† Significantly more women were in university-based programs than in community-based programs ($p = .02$).

distribution of women among each of the three years of training was approximately equal to that of the men residents. There was, however, a significant difference in the distribution of women residents among the three types of training programs. Women residents were significantly more likely to receive their training in university-based programs than in either of the two types of community-based programs.

PERSONAL OR EMOTIONAL PROBLEMS

The residents were asked whether they had experienced personal, emotional, behavioral, or mental problems that concerned them in the preceding year. A significantly higher percentage of women residents than men, 64.1 versus 50.7 percent, reported that they had (chi-square = 5.75, df = 1, $p = .02$). The residents who answered "no" were asked to indicate whether they had experienced no problems at all or no severe problems. Those who responded that they had experienced problems were asked to indicate whether their problems were not severe, moderately severe, or severe enough for them to need professional help. A significantly higher percentage of the women than men (17 versus 10 percent, respectively) considered their problems to be severe enough to require professional help, as shown in Table 2.

The hypothesis that first-year residents have a higher incidence, greater severity, and greater need for professional help with personal or emotional problems than their upper-level colleagues was not supported by the data in the study. These findings were somewhat surprising to the author, since, on the basis of previously cited studies, first-year residents would be more likely to experience such problems and might have a greater need for professional help in dealing with them.

Age was not a significant variable in explaining differences in incidence and severity of problems or need for professional help, nor were type of program or

TABLE 2

Number and Percentage of Incidence and Severity of Personal or Emotional Problems and Need for Professional Help of 466 Family Medicine Residents in 33 Residency Programs in Six States, 1985*

Characteristics	Men		Women	
	No.	Percent	No.	Percent
Presence of personal problems during preceding year				
Yes	181	50.7†	66	64.1†
No	176	49.3	37	35.9
Severity of problems				
No problems at all	64	17.9†	11	10.5†
No severe problems	112	31.4	26	25.4
Moderately severe problems	145	40.7	48	47.1
Severe enough to require professional help	36	10.0†	18	17.0†

* The six states were Missouri, Kansas, Oklahoma, Arkansas, Louisiana, and Texas. The number of responses ranged from 428 to 460 because not all residents answered all questions.

† Significant differences ($p = .05$).

location of the program. Gender remained the only significant variable in explaining variance in the three measures (problems in preceding year, health practices, and demographic information) related to personal or emotional problems.

ALCOHOL USE

Data regarding the residents' use of alcohol—including whether alcohol was ever used and the frequency, quantity, and patterns of its use—are shown in Table 3. A significantly higher percentage of the women residents reported using alcoholic beverages than did men residents: 92 percent of women versus 84.1 percent of men. A higher percentage of the women residents reported daily use of alcohol (4.9 percent) than did their male colleagues (1.9 percent). These gender differences in frequency of alcohol use approached statistical significance ($p = .10$). Analysis of variance revealed that frequency of alcohol use was significantly higher among residents in university-based programs than other programs when the analysis controlled for gender differences and postgraduate year ($F = 3.27$, df = 2; $p = .04$). The women residents, however, were significantly more likely than the men to limit their alcohol intake to one or two drinks per occasion.

The residents were asked whether there had been changes in the pattern of their alcohol use within the preceding two years. The results showed that 41.6 per-

TABLE 3

Number and Percentage of 466 Family Medicine Residents in 33 Residency Programs in Six States with Characteristics of Alcohol Use, 1985*

Characteristics	Men		Women	
	No.	Percent	No.	Percent
Used alcohol within preceding year				
Yes	301	84.1†	92	92.0†
No	57	15.9†	8	8.0†
Frequency of use				
Daily	7	1.9†	5	4.9†
2 to 6 days a week	84	23.1	19	18.6
Once a week or less	215	59.2†	70	68.6†
None	57	15.7†	8	7.8†
Amount of alcohol used per occasion of use				
5 or more drinks	3	1.0	1	1.0
3 to 4 drinks	46	13.0†	8	7.8†
1 to 2 drinks	247	69.9†	85	83.3†
None	57	16.1†	8	7.8†
Pattern of use				
More than two years before	31	8.7†	18	17.6†
Same as two years before	120	33.7†	52	51.0†
Less than two years before	148	41.6†	24	23.5†
None	57	16.0†	8	7.8†

* Number of responses ranged from 455 to 465 because not all residents answered all questions.

† Significant difference ($p = .05$).

cent of the men residents reported using less alcohol than two years before versus only 23.5 percent of the women. The women's pattern of use was somewhat more likely to have remained stable over the preceding two years; however, a significantly higher percentage of the women (17.6 percent) than men (8.7 percent) reported using more alcohol than two years before.

A significant association between the incidence of personal or emotional problems and frequency of alcohol use was demonstrated by two-way analysis of variance when the analysis controlled for gender differences ($F = 2.57$, df = 2; $p = .05$). There was also a significant association between the severity of personal and emotional problems and frequency of alcohol use when the analysis controlled for gender differences in the severity of problems ($F = 4.1$, df = 1; $p = .05$). The age of the resident was not a significant variable in explaining the differences in whether alcohol was used or in the quantity or pattern of use. However, age was a significant variable, when the analysis controlled for gender differences, in explaining differences in the frequency of alcohol use ($F = 3.62$, df = 4/395, $p = .007$): the older residents used alcohol more frequently than their younger colleagues.

SMOKING

The incidence of cigarette smoking was very low in both men and women residents. Less than 5 percent of the residents smoked at the time of the survey: 4.1 percent of men and 4.9 percent of women. These gender differences in smoking approached statistical significance ($p = .11$). More of the women than the men reported they had never smoked (86.4 percent versus 77.4 percent). The women who smoked were significantly older than their nonsmoking women colleagues and older than their male colleagues who smoked ($F = 5.12$, df = 2; $p = .001$). The mean age of the women smokers was 32.6 years versus 28.9 years among women who had never smoked. The mean age of men who smoked was 29.6 years and did not differ significantly from that of the men who had never smoked. A significant association between personal or emotional problems and cigarette smoking was demonstrated by partial correlation coefficients, controlled for both age and gender ($r = .12$, $p = .02$). Although the link between personal or emotional problems and cigarette smoking was identified in this study population, the gender differences in rates of smoking were not significant.

WEIGHT

The residents were asked whether they considered themselves to be overweight, underweight, or about average. These terms were not further defined, and, therefore, the responses represent the residents' own perceptions of ideal personal weight. Significantly more women than men considered themselves to be overweight: 46.6 percent versus 33.6 percent (chi-square = 8.65, df = 2; $p = .001$). A significantly higher percentage of the women than the men reported limiting their calorie intake, 86.6 percent versus 62.8 percent (chi-square = 10.2, df = 1; $p = .001$).

A significant association between personal or emotional problems and efforts to limit calories, when the analysis controlled for gender differences in the incidence of personal problems, was demonstrated by analysis of variance ($F = 8.09$, df = 2; $p = .001$). The women who reported having personal or emotional problems were significantly more likely to make efforts to reduce their calorie intake than the men who reported having such problems.

EXERCISE

The results indicated that approximately 12 percent of the residents believed they got enough exercise. Less than 10 percent considered themselves to be much more active than others their same age and gender. Approximately 3 percent of the respondents reported exercising daily, and approximately 30 percent exercised as frequently as two times a week. The remaining 65 percent reported engaging in aerobic exercise once a week or less. There were no significant gender differences in frequency of exercise. In fact, the frequency with which the men and women residents engaged in aerobic exercise was remarkably similar.

Discussion

Do women family medicine residents experience more personal or emotional problems during their residency training than their male colleagues? The responses from the sample of 466 family medicine residents in the study reported here suggest that they do. Significantly more of these women residents reported they had experienced personal or emotional problems during the preceding year than did their male colleagues. And, although the women reported a higher incidence of problems, both the women and men reported a very high incidence of personal or emotional problems during the preceding year: 64.1 percent of the women and 50.7 percent of the men.

More disconcerting to the investigator than the large percentage who reported having problems were the findings that one in six of the women residents and one in 10 of the men residents considered their problems to be severe enough to need professional help. These proportions represent a significantly greater difference in perceived severity of the problems between women and men residents and is considerably higher for both than would be expected in the general population. National survey data (15) indicated that only 5.2 percent of women and 4.9 percent of men experienced personal or emotional problems they considered severe enough to require professional help. It is recognized that there may be gender differences in help-seeking behavior and that women may be somewhat more likely than men to seek professional help. However, the author of the present study believes that the gender differences found in her study represent a true gender difference in both incidence of personal or emotional problems and the degree of severity of the problems. Women residents function in what remains a male-dominated field. Many have assumed multiple roles as physician, wife, mother, colleague, and other important societal roles. Yet the societal norms of women as care-givers, nurturers, and primary parent often seem to be at cross purposes with the goals of medical education.

The question of whether women family medicine residents are at increased risk for maladaptive coping behaviors for dealing with increased incidence and severity of personal or emotional problems cannot be answered definitively on the basis of the present study. However, the findings do point to some causes for concern. The gender differences in the percentage of residents who used alcohol in the preceding year, the increased use of alcohol by women residents over the two years prior to the survey, and the higher incidence of daily use of alcohol by women than men residents must be taken seriously. The significant association between increased incidence of personal or emotional problems and greater frequency of alcohol use supports the hypothesized link between emotional problems and maladaptive coping behaviors

related to alcohol use. This statistical association between emotional problems and alcohol use was still present when the incidence of personal problems was controlled for gender differences by analysis of variance.

The data from this study demonstrate an association between an increased incidence of personal or emotional problems and cigarette smoking for both genders. Since the differences in rates of smoking by gender did not differ significantly, an association between increased incidence of personal or emotional problems among women residents and increased cigarette smoking by women could not be demonstrated. These findings suggest that residents' smoking behavior may serve as something of a barometer of their personal or emotional problems even though it is not possible to state which precedes the other.

Almost half of the women considered themselves to be overweight whereas only about one-third of the men had similar concerns. The hypothesized relationship between a higher incidence of personal or emotional problems and weight-reducing behavior or continued efforts to reduce their calorie intake by women residents was supported by analysis of variance. The high number of women who reported limiting their calorie intake (80.6 percent) seems to indicate a preoccupation with weight control that exceeds the norm in a general population.

The findings support the conclusion that neither men nor women residents use frequent exercise as a means of dealing with personal or emotional problems. Since there were no significant differences in amount or type of exercise by gender, the hypothesized relationship between an increased incidence of emotional problems and infrequent exercise among women residents was not supported.

The significant findings presented here point out the usefulness of conceptualizing the relationship between stress-related problems among family medicine residents and the potential for maladaptive coping strategies. Identifying effective coping strategies that would not place the individual at increased risk for substance abuse or other health problems should be incorporated into the residency education program. Such efforts require action rather than mere words of encouragement or crisis management of individual problems as they arise.

Assistance from faculty members in reducing erratic work schedules and sleep deprivation would be beneficial, as would increased social support from both faculty members and peers. Reducing the work load by reducing the number of patients seen by residents and shortening the working hours per day to nine or fewer, would be especially useful to women residents. More flexibility in the operation of residency programs, such as providing for longer maternity leaves or shared residency positions, should be offered in an effort to reduce the risk of maladaptive coping behavior. Assertiveness training should be offered to residents to help them develop skills for dealing with patients who place unreasonable demands on their time. This would be especially appropriate for women residents because the acculturation of many women has encouraged submissiveness rather than assertiveness.

Educational efforts to increase the coping skills of residents should be given a higher priority than at present. All residents could benefit from such training. It is critically important to identify those residents at high risk for personal or emotional problems early in their training and assist them in developing effective coping strategies. If such problems go unde-

tected, the risk of the resident becoming an impaired physician are real. Smoking, misuse of alcohol, overeating, and lack of physical exercise are risks to the physician's health that could be prevented or ameliorated by helping residents develop effective ways of managing stress.

References

1. ALEXANDER, D., MONK, J. S., and JONAS, A. P. Occupational Stress, Personal Strain, and Coping Among Residents and Faculty Members. *J. Med. Educ.*, **60**:830–839, 1985.
2. MAY, H. J., and REVICKI, D. A. Professional Stress Among Family Physicians. *J. Fam. Pract.*, **20**:165–171, 1985.
3. FREY, J., DEMICK, J., and BIBACE, R. Variations in Physicians' Feelings of Control During a Family Practice Residency. *J. Med. Educ.*, **56**:50–56, 1981.
4. MAZIE, B. Job Stress, Psychological Health, and Social Support of Family Practice Residents. *J. Med. Educ.*, **60**:935–941, 1985.
5. FORD, C. V., and WENTZ, D. K. The Internship Year: A Study of Sleep, Mood States, and Psychophysiologic Parameters. *South. Med. J.*, **77**:1435–1442, 1984.
6. MCCUE, J. D. The Distress of Internship/Causes and Prevention. *N. Engl. J. Med.*, **312**:449–452, 1985.
7. RUDNER, H. L. Stress in Family Practice Residents. *Can. Fam. Physician*, **32**:319–323, 1986.
8. RUDNER, H. L. Stress and Coping Mechanisms in a Group of Family Practice Residents. *J. Med. Educ.*, **60**:564–566, 1985.
9. SMITH, J. W., DENNY, W. F., and WITZKE, D. B. Emotional Impairment in Internal Medicine House Staff/Results of a National Survey. *J.A.M.A.*, **255**:1155–1158, 1986.
10. LLOYD, C., and GARTRELL, N. K. Sex Differences in Medical Student Mental Health. *Am. J. Psychiatry*, **138**:1346–1351, 1981.
11. HOFEREK, M. J., and SARNOWSKI, A. A. Feelings of Loneliness in Women Students. *J. Med. Educ.*, **56**:397–403, 1981.
12. PFEIFFER, R. J. Early Adult Development in the Medical Student. *Mayo Clin. Proc.*, **58**:127–134, 1983.
13. LORBER, J., and ECKER, M. Career Development of Female and Male Physicians. *J. Med. Educ.*, **58**:447–456, 1983.
14. EISENBERG, C. Women as Physicians. *J. Med. Educ.*, **58**:534–541, 1983.
15. SCHOENBORN, C. A. Basic Data From Wave I of the National Survey of Personal Health Practices and Consequences. *Vital and Health Statistics/Series 15, No. 2.* (DHHS publication No. [PHS] 81-1163). Washington, D.C.: Government Printing Office, 1981.